

Data Collection And Preprocessing Phase

Date	15 July 2024
Team ID	739827
Project Title	Thyroid disease classification using machine learning
Maximum Marks	2 Marks

Data Collection Plan & Raw Data Sources Identification Report

Elevate your data strategy with the Data Collection Plan and the Raw Data Sources report, ensuring meticulous data curation and integrity for informed decision-making in every analysis and decision-making endeavor.

Section	Description
Project Overview	A project on thyroid disease classification using machine learning involves developing a model to accurately predict thyroid conditions based on patient data, improving diagnosis and treatment efficiency.
Data Collection Plan	The data collection plan involves gathering patient records, including symptoms, lab results, and medical history, ensuring a diverse and

	representative dataset for training the machine learning model.
Raw Data Sources Identified	In a thyroid disease classification project using Machine Learning, raw data sources like patient medical records, lab test results, imaging scans, and symptom descriptions are identified and used to train the ML model for accurate classification of thyroid conditions.

Raw Data Sources Report:

Source Name	Description	Location/URL	Format	Size	Access Permission
Kaggle Dataset	The dataset comprises patient details (age, gender) and Thyroid classification outcomes	https://www.kaggle.com/code/dariomandarin/thyroid-classification	CSV	373kb	Public